AD A040911

ARI TECHNICAL REPORT
TR-76-A2

12

Evaluation of the Effectiveness of Training Devices: Validation of the Predictive Model

by

George R. Wheaton, Andrew M. Rose,
Paul W. Fingerman, and Russell L. Leonard
AMERICAN INSTITUTES FOR RESEARCH
1055 Thomas Jefferson Street N.W.
Washington, D.C. 20007

and

G. Gary Boycan
ARMY RESEARCH INSTITUTE FOR THE
BEHAVIORAL AND SOCIAL SCIENCES

OCTOBER 1976

Contract DAHC-19-73-0049
Final report

Prepared for

ari

U.S. ARMY RESEARCH INSTITUTE
for the BEHAVIORAL and SOCIAL SCIENCES
1300 Wilson Boulevard
Arlington, Virginia 22209

Approved for public release; distribution unlimited

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

J. E. UHLANER
Technical Director

W. C. MAUS
COL, GS
Commander

Research accomplished
under contract to the Department of the Army
American Institutes for Research

NOTICES

(18) ARI

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (19) Technical Report TR-76-A2	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) EVALUATION OF THE EFFECTIVENESS OF TRAINING DEVICES: VALIDATION OF THE PREDICTIVE MODEL.		5. TYPE OF REPORT & PERIOD COVERED (9) Final report.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) George R./Wheaton, Andrew M./Rose, Paul W. /Fingerman, Russell L./Leonard, Jr. (AIR), and G. Gary/Boycan (ARI)		8. CONTRACT OR GRANT NUMBER(s) (15) DAHC 19-73-C-0049 new
9. PERFORMING ORGANIZATION NAME AND ADDRESS American Institutes for Research 1055 Thomas Jefferson St., N.W. Washington, D.C. 20007		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS (16) 2Q763731A762
11. CONTROLLING OFFICE NAME AND ADDRESS Deputy Chief of Staff for Operations and Plans U.S. Army Washington, D.C.		12. REPORT DATE (11) October 1976
		13. NUMBER OF PAGES 51
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office) U.S. Army Research Institute for the Behavioral and Social Sciences PERI-OU 1300 Wilson Blvd., Arlington, VA 22209		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(12) 59p.

18. SUPPLEMENTARY NOTES

Research monitored by the Unit Training and Evaluation Systems Technical Area of ARI

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Transfer of training
Training devices
Modeling
Training device evaluation
Training effectiveness research
Armor training
Live firing

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This report describes efforts to develop and validate a transfer of training model which can be used to predict the effectiveness of Army training devices. Initial efforts in accomplishing this objective included: 1) a review of transfer of training research and previous attempts at developing a predictive model; 2) the generation and refinement of a

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

023 450

20. preliminary model; 3) the consideration of methodological issues regarding the assessment of transfer of training; and 4) an evaluation of the effectiveness of representative training devices to assess the feasibility of applying the model and generating indices of device effectiveness.

Upon completion of these developmental activities, two field experiments were performed using training devices from the Armor Branch of Combat Arms. The purpose of these experiments was to provide empirical transfer data against which to validate the model. Experiment 1 assessed the effectiveness of three Burst-on-Target (BOT) training devices for preparing AIT trainees to perform BOT on the 3A102B laser device. While some differences in the devices were noted, all proved to be reasonably effective trainers. Experiment ~~II~~ 2 compared the effectiveness of two devices and three levels of training proficiency for preparing AIT trainees to perform a live-fire tracking task using the main gun of the M60A1 tank. The two devices were not particularly effective when compared to an untrained control group. However, the more highly trained trainees were more accurate than the untrained control group at the end of the live-fire task.

a transfer of training

The predictive model was employed to generate predictions of effectiveness for the training devices used in both experiments. These predictions were then compared to the actual effectiveness data obtained from the field experiments. These comparisons provided support for the model's predictions, and revisions of the model were not warranted on the basis of the comparisons. Some revisions were suggested on other rational grounds. These adjustments are discussed in detail and their underlying rationales are presented.

1473B

FOREWORD

This report summarizes efforts to develop and validate a transfer of training model which could be used to predict the effectiveness of Army training devices. Details of the project have been printed by the Army Research Institute for the Behavioral and Social Sciences (ARI) in Research Memorandums 76-6, 76-16, 76-18, and 76-19. The research was conducted jointly by personnel of ARI's Unit Training and Evaluation Systems Technical Area and the American Institutes for Research under contract DAHC-19-73-C-0049, in response to special requirements of the Army Deputy Chief of Staff for Operations and Plans (DCSOPS) and RDTE Project 2Q763731A762.

J. E. UHLANER
Technical Director

EVALUATION OF THE EFFECTIVENESS OF TRAINING DEVICES:
VALIDATION OF THE PREDICTIVE MODEL

BRIEF

Requirement:

To evaluate a model for predicting training device effectiveness. The model takes into consideration the performance tasks to be trained, the capabilities of trainees who will use the device and the manner in which the device is to be used.

Procedure:

This report reviews efforts to develop a transfer of training model which provides guidelines for generating estimates of device effectiveness and presents the results of initial efforts to validate the model.

Upon completion of initial developmental activities, two field experiments were conducted using training devices from the Armor Branch of Combat Arms. These experiments were designed to obtain empirical transfer data against which to validate the model. Experiment I assesses the effectiveness of three Burst-on-Target (BOT) training devices for preparing trainees to perform BOT on the 3A102B laser device. Experiment II compared the effectiveness of two devices and three levels of training proficiency for preparing trainees to perform a live-fire tracking task using the main gun of the M60A1 tank. The model was used to generate predictions of effectiveness for the training devices used in both experiments. These predictions were then compared to the actual effectiveness data obtained from the field experiments.

Findings:

Applications of the model during the developmental and validation stages demonstrated that it was feasible to develop stable estimates of device effectiveness, provided that some form of task descriptive or task analytic information was available for both the training device and the operational equipment. Comparisons between predicted and actual device effectiveness provided provisional support fot the model's predictive validity. Outcome of the field studies, however, precluded a rigorous test of the model's validity in that neither set of devices was found to differ significantly in actual effectiveness. Lacking such differences in actual effectiveness, validation required only an analogous equivalence among the estimates provided by the model. This equality in predicted values of effectiveness was found for the devices studied in each experiment.

Utilization of Findings:

Application of the model at various stages in the design and development cycle for a device could prove useful in several ways. Estimates of device effectiveness could serve to identify potentially unsatisfactory devices prior to hardware development. Similarly, during the process of device redesign, revised estimates of effectiveness based on proposed changes could aid in deciding which modifications should be implemented and which dropped from consideration. Moreover, should future evaluations of the model provide clearer evidence of its predictive validity, device effectiveness estimates could serve as one basis for selecting among competing devices.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 INTRODUCTION

One of the most complex problems facing Army planners is the design and development of effective training systems, particularly where the use of operational equipment for training purposes is impractical. A number of factors constrain the use of operational equipment in a training role. These include reduced military budgets with consequent reduced availability of actual hardware for training purposes, reduced availability of large-scale training areas and ranges, and finally, growing concern for the ecological damage which can arise from mechanized field training maneuvers.

In order to deal with the limitations resulting from the reduced use of operational hardware the Army has turned increasingly to the use of training devices which simulate the operational situation. These training devices are designed to meet the needs of a variety of students who enter the training situation lacking varying degrees of knowledge or skill. To the extent that exposure to the training device(s) imparts necessary knowledge and facilitates performance at specified criterion levels on an operational task, the training device is judged to be effective.

The training device, of course, is but one component of a larger and more complex training system. The effectiveness of the device, therefore, depends in part on other aspects of the overall system. The development of a training system starts with a statement of the requirement for training (e.g., a new operational system is being developed for which trained operators will eventually be required; alternatively, a training need is identified which is relevant to several operational systems). The next step is to identify what needs to be trained (the skills and knowledges required to man the operational system successfully, which are not possessed by untrained personnel) and to specify the general training system which will supply training in the required areas. At this stage the specification is still a "functional" one, oriented toward the goals and objectives of training as opposed to training hardware.

As planning of the training program progresses, the level of detail increases and the means by which training is to be implemented begins to be explored. Decisions are made regarding classroom and on-the-job

training, the length of the course(s), and the requirements for training devices and training aids to support training. Many of these decisions are fairly straightforward and can be made by experienced training specialists familiar with the personnel resources and needs of the Army. Questions regarding training devices, however, are not as readily amenable to such decision making. How and when to use training devices, how to design them, and what to spend on them are issues which, in the past, have necessarily been dealt with in a fairly arbitrary manner due to the lack of objective bases on which to make such decisions. Sound instructional decisions regarding the use of training devices are more likely to be made when based upon a conceptual framework and objective methodology which can be employed to forecast training device effectiveness in advance of expensive developmental activities.

The primary goal of the present project is the development of a model which can be used to predict and to evaluate the effectiveness of training devices. The modeling is particularly aimed at describing how device design, device use, training approach and individual ability interact to influence device effectiveness. The standards of effectiveness emphasize transfer of military skills from training to the operational setting. Initial modeling efforts resulted in development of a training-content by training-process model in which device effectiveness was viewed as a function of: 1) the potential for transfer, 2) the magnitude of the trainees' learning deficit, and 3) the appropriateness of the training techniques used to overcome that deficit. Subsequent efforts have continued to refine and clarify the model and field experiments have been carried out to assess its validity.

Purpose of the Report

As the summary document in this series, this report describes efforts to validate the predictive model. First, it provides a synopsis of the overall project activities as background to the validation exercise. These included: 1) a literature review and preliminary model development (Wheaton, Rose, Fingerman, Korotkin, & Holding, 1976), 2) and elaboration of the preliminary model, along with a specification of the procedures for its application (Wheaton, Fingerman, Rose, & Leonard, 1976), 3) the conduct of a field experiment designed to determine the effectiveness of devices used to train Burst on Target (BOT) skills in tank gunnery (Wheaton, Rose, Fingerman, Leonard & Boycan, 1976), and 4) the conduct of a field experiment concerned with training of tracking skills in tank gunnery with transfer to a live-fire exercise (Rose, Wheaton, Leonard, Fingerman, & Boycan, 1976).

Following this presentation of background material the report describes how the model was applied to various devices used in the burst-on-target and tracking field studies to generate predictions of transfer of training. These predictions are then compared with the obtained empirical transfer data in order to provide estimates of the model's validity. The remainder of the report discusses the outcome of the validation exercise in terms of suggested revisions in the model and recommendations for further work in the area of device evaluation and transfer of training research.

2.0 BACKGROUND

2.1 Development of Training Device Effectiveness Model

As the initial step in the development and evaluation of a training effectiveness model, a comprehensive survey of the literature was undertaken (Wheaton, Rose, Fingerman, Korotkin, and Holding, 1976). Several different kinds of literature potentially bearing on the prediction of device effectiveness were exhaustively reviewed, reduced, and analyzed. These included: 1) previous methods and models dealing with the design and evaluation of training programs and the prescription or prediction of device effectiveness; 2) major psychological theories of transfer of training together with their implications for a predictive model and for device evaluation; and 3) empirical data dealing with a host of substantive issues, particularly in terms of specific variables and their impact on transfer of training.

In conducting the review, over 2,000 abstracts were screened for possible relevance. Based upon this initial evaluation, over 250 documents of direct relevance to either the structure of the model or to issues surrounding its application were examined in detail. To the best of our knowledge, only one previous review, compiled by Bernstein and Gonzalez (1971) and indexed by Blaiwes and Regan (1970), has covered, to a comparable extent, what has proved to be a very diverse and fragmented literature.

The review effort culminated in the development of a preliminary transfer of training model as shown in Figure 1. This preliminary model incorporated most of the central issues involved in training device effectiveness that were revealed through the analysis of previous models, methods, and empirical data. In particular it dealt with two major classes of variables:

1 those associated with determining whether a training device does, in fact, elicit the behaviors which are required in operational settings; these were termed "appropriateness" variables.

2. those variables associated with actually learning these behaviors; these were called "efficiency" variables.

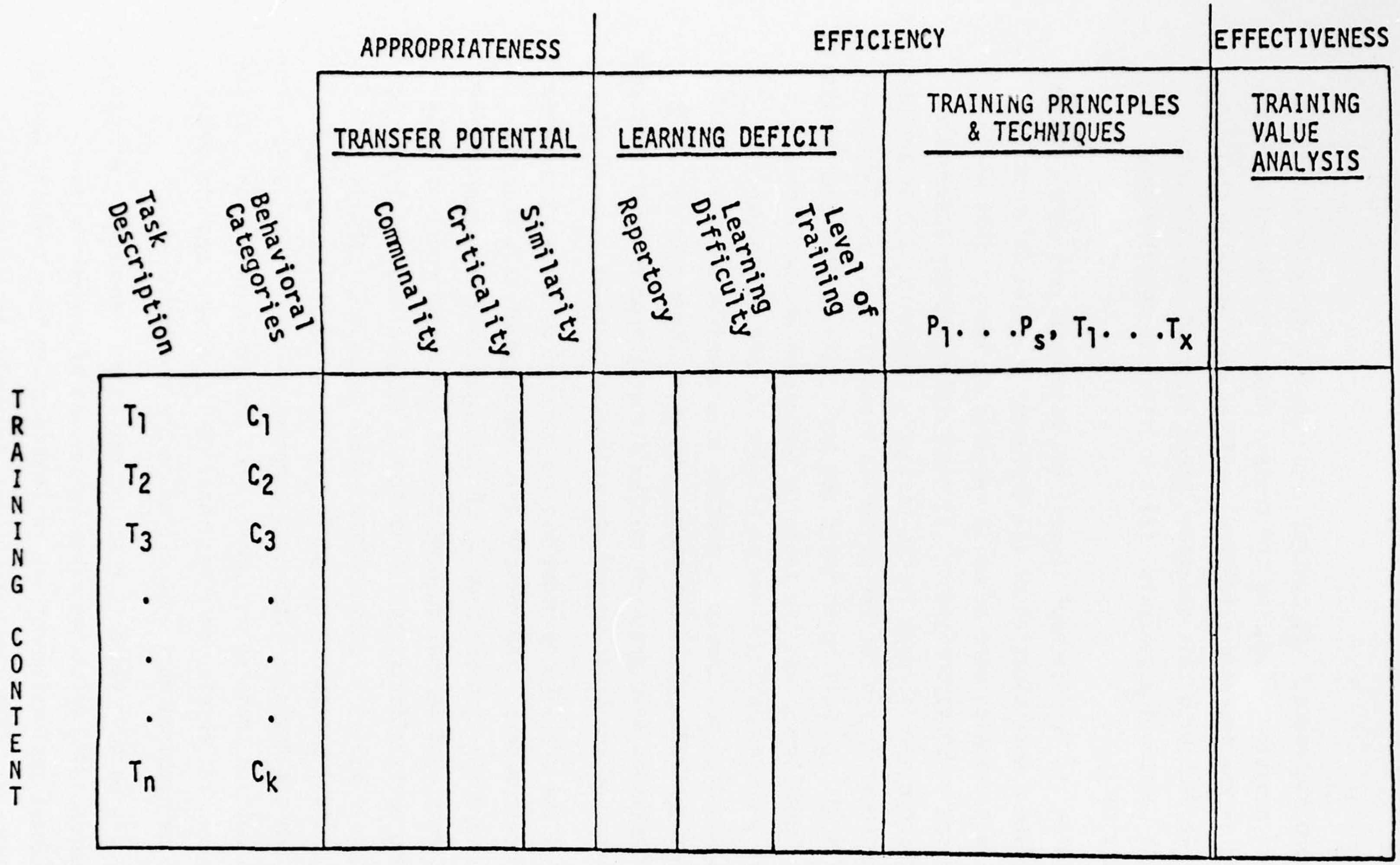

Figure 1. Preliminary structural model.
(From Wheaton, Rose, Fingerman, Korotkin, and Holding, 1976)

Under "appropriateness," the central issue was the transfer potential of the device in question. Assuming the trainee became proficient on the tasks presented in the training situation, would he then meet the training requirements? To deal with this question, three types of analyses were proposed: (1) a communality analysis, (2) a criticality analysis, and (3) a similarity analysis.

In addressing the "efficiency" issues, two major analyses were proposed. The first involved a determination of the trainees' learning deficits: an assessment of what trainees were actually required to learn. This was addressed by three proposed analyses: (1) determining whether appropriate skills and knowledge were already in the trainee's repertory; (2) establishing the proficiency requirements for the criterion transfer task; and (3) estimating how difficult it would be to learn the task. The second major analysis subsumed under "efficiency" was the training techniques and principles analysis. This proposed analysis was an attempt to make direct use of empirical data and training theory as applied to a specific situation in order to predict the efficacy of training. The basic input data for all of these proposed analyses were presumed to stem directly from or to be derivable from task analyses of the training and operational situations.

In summary, the preliminary model was represented as a training-content by training-process matrix. The content axis consisted of task analytic data, while the process axis was made up of two major headings; appropriateness and efficiency, and several subheadings. A functional model was only implied in these early efforts; basically it was assumed that the inputs to the appropriateness and efficiency analyses would be task or subtask descriptions (and the behavioral categories for these tasks) of the operational system and the training situation, combined with a physical description of the operational and training equipment. The precise nature of the measurements to be taken, the results of these individual analyses, and how these measures would be combined were unspecified in the preliminary model.

These details were provided in a subsequent report (Wheaton, Fingerman, Rose, and Leonard, 1976) which described refinement of the preliminary model and an actual application of the refined model to four training devices used in tank gunnery. With minor exceptions, the model in its revised form

retained the basic structure of the preliminary model. While specific decisions regarding the implementation of the various analyses were made, the basic rationale for the general types of analyses remained unchanged. Training device effectiveness was still viewed as a function of the transfer potential for the device, the learning deficit of the trainees, and the extent to which appropriate training techniques were utilized in the device.

As mentioned previously, training effectiveness must be viewed within a system context since effectiveness may be moderated by a host of potent variables external to the device itself, such as device acceptance, other instructional support, etc. While it was still felt that many of these variables could be considered more appropriately in a training system effectiveness model, provision was made for an extended application of the current device effectiveness model. For instance, it was found that the type and amount of supporting classroom instruction could be incorporated into the learning deficit portion of the model.

Figure 2 presents the refined structural and functional model. The structural model is described under three major headings: Inputs, Processes, and Outputs. Functional relationships are indicated by arrows leading from inputs through processes to outputs. As in the preliminary model, the three major analyses to be performed in applying the model are transfer potential analysis, learning deficit analysis, and training techniques analysis. Details of the procedures for applying the model are presented in a previous report (Wheaton, et al., 1976).

While an abridged version is given in Appendix 1, this Appendix omits a discussion of the rationale for the specific analyses and computational procedures and is designed as a self-contained user's manual.

Using these procedures an experimental application of the model was undertaken. The purpose of this preliminary application was twofold. First, it was designed to assess the feasibility of applying the model. Feasibility included an assessment of the time in a training device's "life cycle" at which the model could be applied, either at the Training Device Requirement (TDR) stage, or at the prototypic device stage. Another aspect of feasibility concerned the potential application of the model for system versus nonsystem training devices. A final aspect of feasibility was consideration of the

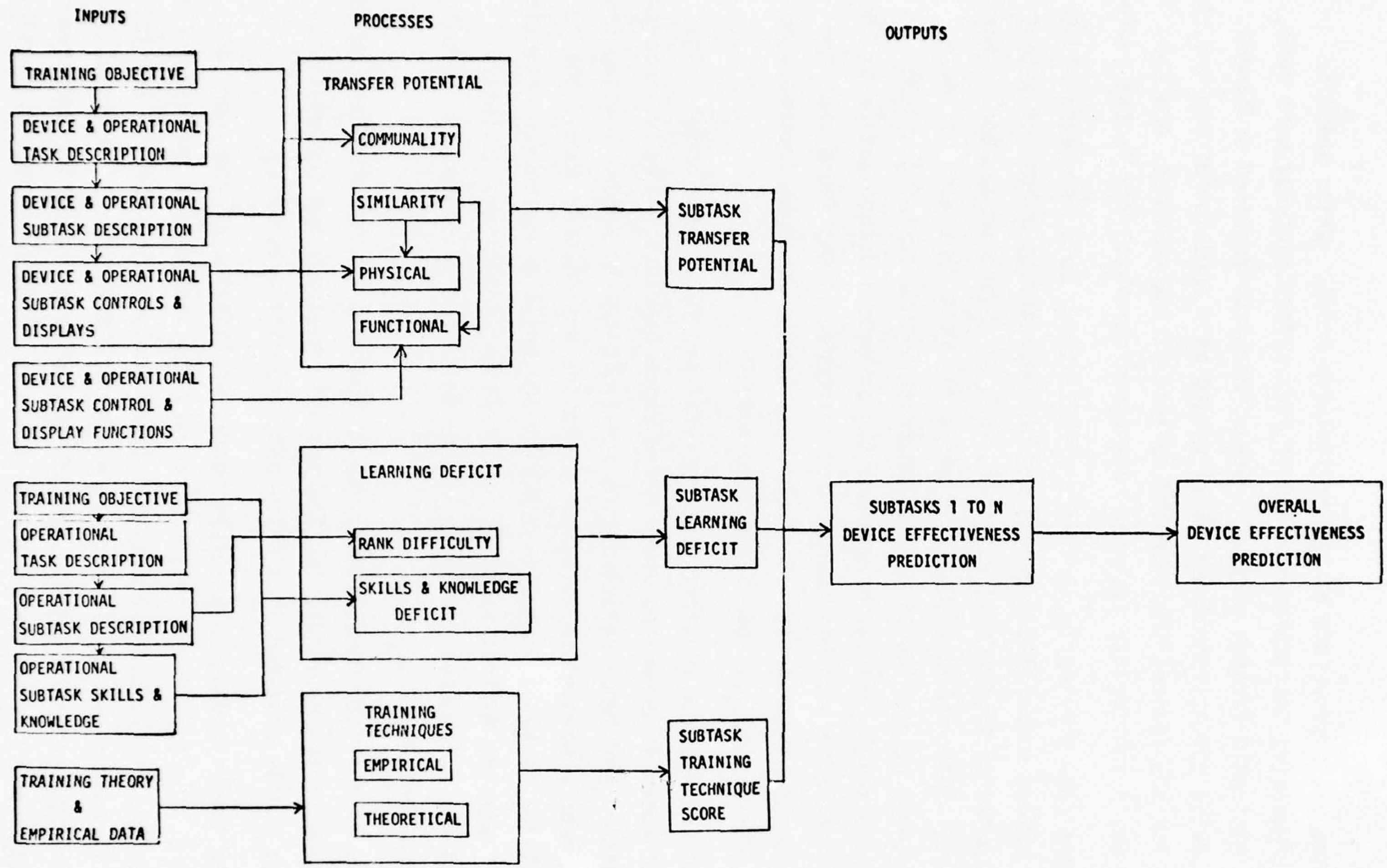

Figure 2. Structural and functional model of training device effectiveness.
(From Wheaton, Fingerman, Rose, and Leonard, 1976)

"omniscience" required of the analyst--whether the analyses could be performed without excessively subjective judgments. The second purpose of the application was to determine the reliability of the procedures. In most cases, modeling data were generated independently by four senior project staff members permitting computation of reliability indices. Two criterion tasks were examined including: (1) firing the main gun of the M60A1 tank using the M-32 sight, and (2) applying burst-on-target (BOT) adjustment of fire using the M-32 sight and the "standard" procedures. Four devices were evaluated: the 17-4 Burst-on-Target Trainer, the 17-B4 (Wiley) Conduct-of-Fire Trainer, the M-55 Conduct-of-Fire Trainer, and SIMFIRE. These devices were selected for two reasons. First, each could conceivably be used to train both criterion tasks. Second, it was assumed that this set of devices would yield at least modest variation in obtained values for different model parameters. Thus, the impact of these variations on the model's feasibility and reliability across a range of predictions could be assessed.

Results of the preliminary application were favorable. The feasibility of applying the model to a variety of training situations was demonstrated and it appeared that values could be derived for relevant parameters with reasonable reliability. Accordingly, detailed consideration was given to an evaluation of the validity of the model's forecasts of device effectiveness. This involved the generation of a research plan for the conduct of two field studies.

2.2 Conduct of Field Studies

Planning for the validation exercise involved resolution of a number of methodological issues. These included, for instance, selecting predictive rather than construct validity as the basis upon which to initially "evaluate" the model, choosing the experimental design, and specifying the measure of transfer which would be most appropriate to the goals of the project. Based upon these considerations, two field experiments were conceived which could provide empirical data against which predictions from the model could be evaluated.

Both experiments were formulated after a review of current Army training devices and school curricula, and interviews with cognizant personnel. In each case the major consideration in selecting devices was that those chosen all addressed a common training objective in order to facilitate comparisons. Further constraints included local availability, feasibility of use, and ease of instructor and trainee orientation and operation. Based on these considerations, two sets of devices were selected. For the first experiment the decision was made to study devices potentially or actually used to train Burst-on-Target (BOT) adjustment of fire within the Armor Branch of Combat Arms. The second experiment addressed devices within the Armor Branch which provided training in tracking and firing at moving targets.

Specifically, the first field experiment compared the effectiveness of three training devices for preparing Advanced Individual Training (AIT) personnel to apply BOT techniques with the 3A102B laser device mounted in the M60A1 tank. The three devices were: (1) the 17-4 BOT trainer (the "Green Hornet"), (2) a modified version of the 17-4 trainer which was fabricated specifically for this experiment by the Training Aids Department at Fort Knox, and (3) the 17-B4 (Wiley) Conduct-of-Fire Trainer.

Three groups of 20 trainees each were trained on the devices until they achieved a proficiency criterion of 90% BOT hits in a series of shots, or until they had fired a total of 320 shots. All trainees reached proficiency within the allotted 320-shot period. During training, data were collected on both time (between first and second shots of a BOT engagement) and accuracy measures. Following training, the three experimental groups were transferred to the 3A102B device where each trainees received 80 BOT trials. A fourth group of 20 trainees practiced for 80 shots only on the 3A102B device. This group served as the control group. Time and accuracy data were recorded In addition, device acceptance data were obtained from both trainees and instructors. Details of this first experiment have been reported elsewhere (see Wheaton, Rose, Fingerman, Leonard, and Boycan, 1976).

In the second field study AIT personnel were trained to different levels of proficiency in tracking on two training devices and transferred to a live-fire tracking task using the main gun of the M60A1 tank (see Rose, Wheaton,

Leonard, Fingerman, & Boycan, 1975). The devices selected for evaluation in this experiment were: (1) the M73 coaxial machine gun mounted in the M60A1 tank and firing in the single shot mode; and (2) the 3A102B laser device mounted in the M60A1 tank.

Three groups of 22 trainees were trained on each device until they achieved a proficiency criterion of 30%, 50%, or 70% hits. A seventh group of 22 trainees received no device training and served as the baseline against which to evaluate the other groups. Following training all groups were transferred to a live-fire range where each trainee fired 12 main-gun rounds at a moving tank silhouette target. Time and accuracy data were recorded as were various kinds of miss data.

2.3 Application of Model

Predictions of effectiveness were obtained for the training devices used in each field study by applying the model in accordance with procedures described in Appendix 1. Four major steps were involved. In the first step the transfer potential of the devices was determined. This involved an assessment of the communality (C) of subtasks between each training device and the criterion situation, as well as the similarity (S) between training device and criterion equipment displays and controls. The second step required estimation of a learning deficit on each subtask for the actual trainees who were to serve as subjects in the experiments. The deficits were then weighted (WLD) by the estimated difficulty of learning to perform the various subtasks. In the third step an assessment was made of the extent to which the training devices adhered to various principles of training and/or incorporated useful training techniques (T). The fourth step involved the combination of data from the preceding steps into an index reflecting each device's effectiveness.

To illustrate the first three steps, data are presented in Appendix 1 for the three BOT devices evaluated during the first field study. It should be noted that several of the reported values differ from those presented in previous reports in this series (i.e., Wheaton, Fingerman, Rose, and Leonard, 1976; Wheaton, Rose, Fingerman, Leonard, and Boycan, 1976). These

differences arise since a different criterion task has been assumed in the present effort--namely, applying BOT with the 3A102B laser under the specific conditions of the field study rather than application of BOT with the M60A1 main gun.

The estimates of effectiveness (τ) are obtained by evaluating the equation:

$$\tau = \frac{\sum_{i=1}^{n} C_i S_i T_i WLD_i}{\sum_{i=1}^{n} WLD_i} .$$

This equation was derived and discussed in a previous report (Wheaton et al, 1976) and is intended to represent an estimate of the percentage of transfer of training as defined by Gagné, Foster, and Crowley (1948). The components of the equation refer to the indices obtained from the steps discussed above.

The next section of the report presents the effectiveness estimates obtained for the various devices. In addition, the empirical findings from the two studies are reviewed, and the predicted effectiveness is compared in each case with the obtained transfer data.

3.0 VALIDATION OF MODEL

In order to validate the training effectiveness model, transfer of training values obtained from the BOT field study were compared to predictions of effectiveness provided by the model. Based on the degree of correspondence between obtained and predicted values, revision of the model was considered. The model was then used to generate predictions of effectiveness in the tracking study. The predicted effectiveness values were compared to the transfer values actually obtained from that study. The results from each of these steps are presented in the following sections.

3.1 Predictive Validity--BOT Study

Since the results of the BOT experiment have been described in detail elsewhere (Wheaton et al., 1976) only the transfer effects will be reviewed here. First, no significant difference was found between the three training device groups (i.e., 17-4, 17-4M, and 17-B4) and the control group on BOT accuracy in the transfer task. The training device and control groups initially hit approximately 42% of their shots, and improved to about 60% after 80 transfer trials. While some device groups performed significantly better than the control in one or two particular blocks of 10 transfer trials this was a small and transient effect. The most appropriate characterization of accuracy of BOT performance during transfer was that prior training did not have a pronounced effect.

Prior training did impact on the time-between-shots measure of BOT performance during transfer. All three training device groups were significantly faster than the control group through the first block of 10 transfer trials, while the device groups themselves did not differ from one another. Through 30 trials, all three trained groups (pooled) maintained their superiority, while two of them (pooled 17-4M and 17-B4 groups) remained faster than the control group through 40 trials. Thus, training produced faster performance, regardless of the device used, through 30 trials; after 50 trials, the control students had had sufficient practice to overcome their initial disadvantage.

To determine the amount of transfer of training which occurred, the BOT accuracy and time-between-shots data were couched in terms of "percent transfer" as defined by Gagne, Foster, and Crowley (1948). Three parameters are required to compute percent transfer. For accuracy (or other proficiency scores where improvement in performance results in an increasing score), percent transfer is defined as:

$$\frac{T_n - C}{S - C} \times 100$$

where T_n = the score of a training device group on the nth transfer trial or block of trials,

C = the score of the control group on the first transfer trial or block of trials,

S = the optimal or asymptotic control group score.

If the scores are errors, time, or trials (where improvement in performance results in a decreasing score), percent transfer is defined as

$$\frac{C - T_n}{C - S} \times 100.$$

While C and T_n are readily estimated by data obtained from the BOT experiment, S is more difficult to establish. Since it is usually impractical to establish a "true" asymptotic or optimal score for a control group, a theoretical value is often assumed. In terms of proportion of hits, the theoretically optimum score is 1.0, while a more practical asymptotic value might be .95. For time between shots, an ideal value would be 0 seconds, while a practical asymptote might be 5 seconds.

Percent transfer values for proportion of hits and time between shots using the idealized estimates of S are presented in Table 1. An additional figure-of-merit (FOM) measure is also presented. This number is defined as time per hit (i.e., time between shots divided by proportion of hits) and is a rough index of a trainee's overall performance. A transfer value is presented for each device for initial transfer (T_1, first block of 10 trials) as well as for final transfer (T_8, last block of 10 trials). To provide a baseline for assessing the final values, a "transfer value" is also shown for the control group, using as T_8 the control group performance

TABLE 1

PREDICTED AND OBTAINED TRANSFER IN THE BOT EXPERIMENT

(OPTIMUM S VALUES ASSUMED)

Measure of Performance	Device/ Group	Predicted $\underline{\tau}$	Obtained Transfer Initial Block	Final Block	Assumed S Value
Accuracy					1.0
	17-4	.27	-.02	.32	
	17-4M	.28	.17	.43	
	17-B4	.32	.13	.37	
	Control			.28	
Time					0
	17-4	.27	.18	.39	
	17-4M	.28	.32	.47	
	17-B4	.32	.31	.37	
	Control			.30	
FOM					0/1
	17-4	.27	.15	.59	
	17-4M	.28	.47	.69	
	17-B4	.32	.43	.61	
	Control			.52	

on the eighth block of 10 trials. Table 2 presents analogous transfer values based upon the more practical estimates of S. The only difference between these two tables lies in the magnitude of the obtained transfer values and not in the relative standings of the devices. In discussing the obtained transfer results, attention will be focused on Table 1 rather than Table 2, since in the derivation of predicted values from the model, an ideal estimate of S is always used.

Since in the analysis of accuracy of BOT performance it was shown that no device group performed particularly better than the control group, it may be inferred that the obtained initial transfer values for accuracy in Table 1 reflect little or no transfer. The final BOT accuracy transfer values simply reflect the fact that all groups (including the control group) are more accurate after 80 transfer trials than the control group initially was. The superiority of the training device groups in terms of time is reflected in the moderate positive values of the transfer indices for time data during initial transfer. Since all groups were faster at the end of training, the final transfer values are correspondingly somewhat higher. Finally, on the FOM measure, all three training groups show moderate positive values upon initial transfer, and higher values after 80 trials.

As previously described, predicted effectiveness values ($\underline{\tau}$) were derived by applying the model to the three training devices used in the BOT study. The predicted $\underline{\tau}$s were .27 for the 17-4, .28 for the 17-4M, and .32 for the 17-B4 devices. Since $\underline{\tau}$, as it is currently computed, lies between 0 and +1, these moderate, positive values would seem to differ little, if at all, from one another in practical significance. In other words, the same degree of effectiveness is predicted for all three devices.

Since $\underline{\tau}$s were designed to predict training device effectiveness in terms of transfer of training, they could be compared to the Gagné-Foster-Crowley values obtained in the BOT experiment. Rigorous comparison of $\underline{\tau}$s to the Gagné-Foster-Crowley values is made difficult, however, since neither index has known distributional properties (e.g., variance). Nevertheless, the following may be concluded. First, the model predicted no difference in effectiveness among the three devices. This prediction is consistent with

Table 2

PREDICTED AND OBTAINED TRANSFER IN THE BOT EXPERIMENT

(PRACTICAL S VALUES ASSUMED)

Measure of Performance	Device/ Group	Predicted $\underline{\tau}$	Obtained Transfer Initial Block	Obtained Transfer Final Block	Assumed S Value
Accuracy					.95
	17-4	.27	-.02	.35	
	17-4M	.28	.18	.47	
	17-B4	.32	.14	.41	
	Control			.30	
Time					5
	17-4	.27	.33	.71	
	17-4M	.28	.59	.86	
	17-B4	.32	.58	.69	
	Control			.56	
FOM					5/.95
	17-4	.27	.19	.72	
	17-4M	.28	.58	.84	
	17-B4	.32	.52	.74	
	Control			.64	

the equivalence in transfer actually found among the three devices for all three measures of performance. Second, while the model provides one estimate for each device, the data suggest two different empirical transfer values: moderately positive for the time or the figure-of-merit measures, and zero or near-zero for accuracy. Estimates provided by the model appear to be of the same order of magnitude as the initial transfer values actually obtained for the time and FOM measures. These same estimates seem too high, however, for the degree of transfer obtained on the accuracy measure.

3.2 Consideration of Model Revisions

The outcome of the BOT study was that the three device groups did not differ significantly in performance during transfer, thus yielding Gagne-Foster-Crowley estimates of transfer which did not differ. In order to validate the model, therefore, predictions of transfer effectiveness for the devices should not differ. In fact, the predicted $\underline{\tau}$s do appear to be relatively equivalent, providing some support for the validity of the model. Stronger tests of validity were not available due to the outcome of the experimental study. Had the device groups differed significantly in transfer, the empirical ordering of groups could have been compared to the predicted ordering based on $\underline{\tau}$. Given the fact that the device groups did not differ, no such tests could be made. As a consequence, one had to judge only whether the $\underline{\tau}$s were equivalent. Based on their experience in deriving $\underline{\tau}$, the project team felt that this judgment was appropriate.

Additional support for the model may be indicated by the direction and absolute magnitude of the predicted $\underline{\tau}$s. It seems reasonable to conclude that the model predicted the time and FOM transfer values correctly with respect to direction of transfer and roughly with respect to the magnitude (i.e., small positive) of the effect. There is a problem, however, with respect to the accuracy data. The training devices did not promote positive transfer. The predicted values, therefore, should also reflect zero transfer, but this conclusion is incompatible with that drawn above from the time data. The model did not

generate different predictions for speed and accuracy, since the "training objective" (i.e., the transfer task) was not specified differentially. That is, trainees were instructed to perform the task as quickly and as accurately as possible. One interpretation of this instruction is that the "correct" dependent variable should be the figure of merit (FOM) which assumes that trainees optimize their speed-accuracy trade-off (see Pew, 1969). Given this interpretation, the devices produced positive transfer and the predictions match the obtained results. If the training objective had been stated in terms of either a speed or an accuracy criterion, this information might have been used in making other predictions from the model.

Another discrepancy between predicted and obtained results is that the predicted values do not correspond with "final" transfer in Tables 1 and 2. The explanation for this lack of agreement is simply that the model was not designed to estimate later transfer. In order to incorporate this type of prediction, a major revision in the model would be necessary. The Gagne, Foster, and Crowley parameters that would have to be estimated are T_n (performance of the training group after n trials of the transfer task), and S (the asymptotic performance of the control group). Revisions of this sort would be unproductive at present since the available empirical data are not diverse enough to test the validity of such a revision.

Based on a consideration of the "fit" between the model's predictions of training device effectiveness and the obtained transfer data, there was no compelling reason to modify or revise any of the model's parameters or the computational formulae for $\underline{\tau}$. Therefore, the next section presents the results of the application of the model to the tracking study, using the same procedures for generating predictions. These predictions are then compared with obtained transfer scores, and the results of this comparison are discussed.

3.3 Predictive Validity Tracking Study

Since results of the tracking experiment have been detailed elsewhere (Rose et al., 1976), only the results of the transfer phase need be reviewed here. First, neither the training device on which students practiced (i.e., M73 coax or 3A102B laser) nor the criterion level to which they trained

had an effect on accuracy of performance during transfer. Further, no differences were found in the initial block of three transfer trials between any of the trained groups and the control group. During the final block of three transfer trials, both the 70% laser and 70% coax groups were significantly better than the control group. All groups improved in accuracy over the four blocks of transfer trials.

Empirical transfer values were again derived from these data using the Gagne, Foster, and Crowley (1948) formula. These values are presented in Table 3. The optimum proportion of hits (S) was set at 1.0, and the 30% and 50% criterion groups were pooled within each device, since their performance never differed. The obtained values for initial transfer are all essentially zero, reflecting the fact that the trained and control groups performed initially at the same level. The moderate values for the two 70% groups on final transfer reflect their significant improvement over trials; the lower obtained values for the other two device groups and the control group reflect a similar but smaller improvement.

As previously described, the model was used to derive an estimate of training effectiveness for each device.* The model estimate of $\underline{\tau}$ for both the coax and the laser was found to be $\underline{\tau} = .20$. The fact that the two values of $\underline{\tau}$ are equal is consistent with the equivalence of transfer values obtained between devices. However, the obtained initial transfer values are somewhat lower than the predicted values.

This finding may indicate a problem in defining initial transfer. The performance data and anecdotal evidence suggest that accuracy of the first several main gun rounds which an AIT trainee fires are influenced

---

* It should be noted that only one estimate of $\underline{\tau}$ was derived for each device, while transfer data were available on each device for the three training criterion groups. A basic assumption of the model in evaluating or predicting the effectiveness of a device is that sufficient training will be provided to allow the student to learn what the device can teach (i.e., practice the tasks in common). This "reasonable amount of training" is assumed so that the characteristics of the device can be evaluated independently of other training system variables, such as amount of training. From this point of view, the data from the two 30% and two 50% groups might even be excluded from consideration in evaluating the present model since they may violate the assumption of "reasonable amount of training."

TABLE 3

TRACKING: PREDICTED AND EMPIRICALLY OBTAINED TRANSFER VALUES

Device/ Group	Training Criterion	Predicted $\underline{\tau}$	Obtained Transfer Initial (Trials 1-3)	Final (Trials 9-12)
Coax		.20		
	30%/50%		-.04	.29
	70%		-.08	.44
Laser		.20		
	30%/50%		.08	.19
	70%		-.02	.44
Control				.10

more by apprehension (due to noise, concussion, etc.) than by the effects of training. Thus, it may be unreasonable to expect training effects to show up in transfer performance until this initial apprehension has been overcome. The transfer data suggest that this adaptation occurs late in the series of 12 rounds which subjects fired. For example, the accuracy of two 70% groups did become significantly superior to the control group during the final block of three transfer trials (trials 9 to 12). In conclusion, therefore, the initial transfer effects due to training devices cannot be accurately assessed since they were obscured to an unknown degree by other factors.

4.0 DISCUSSION

4.1 Evaluation of the Current Training Device Effectiveness Model

An overall evaluation of the training device effectiveness model entails consideration of three aspects of model application--feasibility, reliability, and validity. In a sense, each of these represents a criterion which has to be met before one can proceed to the next consideration.

The feasibility of applying the model was demonstrated and discussed in an earlier report in this series (Wheaton et al., 1976). Two concerns were voiced. The first and key issue in practical applications of the model was the feasibility of acquiring the input data which were required. The second issue was whether the procedures required in order to process the input data were reasonably rigorous and objective without being unduly time consuming. Application of the model to a variety of training devices suggested that it was feasible to develop estimates of device effectiveness provided that some form of task-descriptive or task-analytic information (the basic input data) was available for both the training device and the operational situation in which transfer of training would likely occur. Given such information, the procedures designed to process it into a form compatible with components of the model (i.e., communality, similarity, learning deficit, and training technique analyses) were generally precise and efficient.

During these same preliminary applications of the model, assessments were made of the reliability with which the various analyses could be conducted. As indicated earlier, each major analysis was performed independently by four senior project staff members and their individual judgments were compared and contrasted. In general, agreement among the analysts was excellent. While complete agreement was not always achieved, it was possible to arrive at a consensual judgment (i.e., three out of four analysts agree) which yielded a stable estimate in which the judges had confidence.

Considered jointly, results of feasibility and reliability evaluations suggest two restrictions in application of the model. First, because of

its dependence upon reasonably detailed task-descriptive information, its application is limited to devices which are at least at some intermediate stage of design or development. Application of the model at a very early point in the design and development cycle would be difficult. For example, Training Device Requirement (TDR) statements do not currently incorporate detailed task-descriptive information. Second, the various analytic steps require a certain degree of expertise on the part of judges. Consequently, analysts using the model should be experienced with respect to task-descriptive and task-analytic procedures, with the constructs and analyses comprising information theory, as well as with training theory and technology. Even having such experience, judges should probably work in small teams in order to develop estimates based on a consensus of opinion.

Having demonstrated the feasibility and reliability of applying the model, the validity of the model could be addressed. As discussed in an earlier report (Wheaton et al., 1976), two kinds of validity are relevant. The first, predictive validity, addresses the issue of whether the model's output (i.e., an estimate of a training device's effectiveness) can, in fact, predict the effectiveness of different devices. The second, construct validity, deals with the extent to which constructs or parameters of the model hypothesized to influence device effectiveness actually do so. In essence, therefore, predictive validity is concerned with the utility of the model's output while construct validity addresses the model's theoretical structure. In the present project, predictive validity was of more immediate concern than construct validity since an attempt had been made to build the latter into the model. Consequently, the field studies which were conducted focused on whether the model's predictions of relative effectiveness were in agreement with the obtained estimates of transfer of training.

In retrospect, neither field study provided for a truly rigorous test of the model's predictive validity, since neither effort revealed large differences in actual effectiveness among devices. In fact, had this outcome been foreseen, the field experiments would have been modified, to the extent possible, in order to create differences in transfer attributable to practice on the competing devices. Lacking any actual differences

in effectiveness among devices, validation required an analogous equivalence among the estimates provided by the model. In essence, this equality of predicted values of effectiveness was found for the devices studied in each experiment.

Further support for the validity of the model's estimates stems from a comparison of findings from the two field studies. Specifically, the empirically obtained values for initial transfer in the tracking study were smaller than the initial transfer values obtained in the BOT experiment. The predicted effectiveness values behaved similarly; they were lower in the tracking study than in the BOT study.

While these evaluations are encouraging, additional validity studies are certainly warranted. In these efforts both good and poor devices must be used to provide significantly different levels of transfer. The question then would be whether estimates derived from the model track such differences or continue to exhibit only minor variations. Clear evidence of the former case is required before one can claim predictive validity for the model, and before one can begin to interpret the model's scale properties.

4.2 Revisions of the Training Device Effectiveness Model

As already noted, revisions of the effectiveness model were unwarranted given the outcomes of the two validation efforts. Nevertheless, the nature of the confirmatory evidence (i.e., predicting no differences and obtaining no differences) was not very powerful. Therefore, careful consideration was given to revisions of the model based on trends in the data or upon other logical grounds.

For example, although the ordering of transfer for the three BOT devices was not an empirical issue (since they did not differentially impact on transfer), the predicted and obtained orders were not in agreement. Despite the fact that predicted $\underline{\tau}$s were not very different, the 17-B4 was rated as slightly "better" than the other two devices, and the 17-4 and 17-4M devices were rated as equivalent. The trend in the empirical results was that the 17-B4 and 17-4M were equivalent and both were better than the 17-4 device. Part of the explanation for this trend in the empirical results

(discussed in Wheaton et al., 1976) was that the cadillac controls in the 17-B4 device, although physically identical to the controls of the M60A1 tank used during transfer, functioned somewhat differently from those controls (i.e., the dynamics of the two tracking systems differed). This difference was hypothesized to have had a detrimental impact on transfer, causing the 17-B4 to be less effective.

In terms of the model's predictions, this "negative" influence appears in the Similarity analysis; the 17-B4 is given a non-optimal score for functional similarity for the cadillac controls. However, this lowered score does not have the negative impact one would like to see on overall predicted effectiveness; the discrepancy between high physical and low functional similarity apparently is not "weighted" heavily enough. Currently, the procedure for obtaining an overall similarity index is to compute the mean of the functional and the physical similarity for each subtask. An alternative is to penalize a device that has high physical similarity but low functional similarity. This can be accomplished by incorporating a scoring rule for determining similarity, which assigns a negative value to the mean whenever physical similarity exceeds functional similarity. This notion is consistent with the empirical transfer literature concerning stimulus and response similarity. On the response side (parallel to functional similarity) there is ample evidence that different responses in the training and transfer tasks do not in themselves lead to negative transfer. When negative transfer is found, it tends to occur where the responses are highly <u>similar</u>, but differ in small but important ways. An example is afforded by the negative transfer from upward to downward lever movement studied by Adams (1954). Likewise, the traditional Osgood (1949) transfer surface predicts negative transfer when the stimuli are identical (physical similarity) and the responses are antagonistic (functionally dissimilar).

An explanation for why the 17-4 produced less positive transfer than the 17-4M or 17-B4 may be attributable to "user acceptance" (Wheaton, Rose, Fingerman, Leonard, & Boycan, 1976). The user acceptance data for the three devices taken from instructors after they had finished training students showed that the 17-4 received the lowest rating (mean = 31), the 17-B4 received a higher rating (mean = 35.5), and

the 17-4M was rated the highest (mean = 38). These acceptance data reflect the empirical trend of the obtained transfer data, especially in terms of the 17-4M being rated superior to the 17-4. This correspondence suggests that user acceptance has an important influence on training device effectiveness and hence should be incorporated into the model in order to improve the accuracy of predictions.

However, current knowledge about determinants of user acceptance and its impact on transfer is very sketchy. It should be noted that the user acceptance scores that "fit" the data were obtained after the instructors had trained students. Other acceptance data, taken before training, did not fit. Since it would be impossible to obtain the post-training acceptance data before a device is built, it is difficult to imagine how this type of user acceptance should be incorporated. However, this is not meant to dismiss the potential importance of user acceptance. As more definitive knowledge is acquired concerning the impact of user acceptance, the model should certainly be revised to accommodate this information.

Other considerations raised as a result of the comparison between obtained and predicted transfer estimates address issues more fundamental to the structure of the model. These expansions of the model are discussed in the following section.

4.3 Expansion of the Training Device Effectiveness Model

During development and evaluation of the model, a number of training effectiveness and transfer issues arose, the resolution of which were not deemed immediately necessary. This tabling of issues resulted in the efficient development of a provisionally valid model, but one having certain limitations. In the future, these issues should be dealt with and included in expanded versions of the model. Such issues include the prediction of later transfer (i.e., performance after some experience with the operational equipment), the prediction of savings-type transfer measures, and the use of the model as a prescriptive tool (e.g., as a guide to improved device design and/or device modification). These issues and their implications for model development are discussed below.

A question of interest and importance to Army training is how much a given training device will benefit initial (i.e., first exposure) performance on the operational (transfer) equipment. Another important question is how much a given training device will benefit performance after some experience on the operational gear--later transfer. Transfer formulae have been proposed which deal alternatively with initial or later transfer (cf., Hammerton, 1967). The formula adopted for the present version of the model addressed initial transfer. This formula (Gagne, Foster, & Crowley, 1948) compares initial performance of the control group (C) to initial performance of trained groups (T_i). This estimate of initial transfer is defined as:

$$\frac{T_i - C}{S - C}.$$

(S is defined as the optimal or asymptotic performance of the control group.) Alternatively, one may examine later transfer by comparing the performance of the control group to the performance of a trained group after the trained group has practiced on the transfer task for some number of trials (T_n). In this case, transfer is defined as:

$$\frac{T_n - C}{S - C}.$$

As the model now stands, T_n cannot be readily estimated. In order to predict later transfer, one would have to add elements to the model to estimate what acquisition would occur during n additional trials of transfer on the operational equipment. These additional elements would be concerned with: 1) what the trainee was required to learn; 2) what remained for him to learn after training device experience; and 3) how fast he would learn it on the operational (transfer) equipment.

Another important issue for Army training is the savings in terms of time or trials required to reach a specified criterion level of operational proficiency when training devices are employed. While post-hoc measures of savings are available, prediction of savings is incredibly complex. For example, two of the more common savings measures have been discussed by Roscoe (1971, 1972). (Incidentally, all of the savings measures were developed in the context of very high-fidelity simulators in pilot training, where substantial acquisition of skill on the operational equipment has

always been considered necessary. In those settings _initial_ transfer is an interesting but relatively irrelevant consideration.) One of these measures defines savings in terms of the expression:

$$\frac{n - r}{n}$$

where n = the time, trials, or errors for an untrained control group to reach a specified performance criterion on the transfer (or operational) task; and

r = the time, trials, or errors for a trained group to reach the same criterion after transfer from the device to the operational task.

To predict the two parameters, n and r, a model would have to be sensitive to the nature of acquisition in the transfer task. Both the _initial performance level_ upon transfer and the _rate of acquisition_ on the transfer task would have to be estimated in order to make these predictions. If one assumes that acquisition rate is linear, and furthermore is independent of initial performance level, the current model could conceivably provide these predictions. However, such simplifying assumptions are contrary to the results found in most learning are transfer studies. An expanded model would have to deal with these issues before savings measures could be predicted.

The second savings measure explicated by Roscoe (op. cit.) is defined by two parallel formulae:

$$\text{CTEF} = \frac{n - r_x}{x} \qquad \text{and ITEF} = \frac{r_{x-\Delta x} - r_x}{\Delta x}$$

where n is defined as above;

r_x is the time, trials, or errors of the trained group on the _transfer_ task after x time, trials, or errors by that group on the _training_ device.

CTEF is the "cumulative transfer effectiveness function," and ITEF is the "incremental transfer effectiveness function," the latter being measured over increments (Δx) of device training before transfer. To deal with this savings measure of transfer, the model would have to be expanded to treat not only the rate of acquisition on the transfer task, but also the rate

of acquisition during training.

Finally, the Army's concern with training device effectiveness can be viewed in terms of two separate but related questions: 1) how effective will an existing training device be, and 2) how can a training device be constructed or modified to be more effective? While the current model has been specifically designed to address the first of these questions, it could potentially be expanded to address the second "prescriptive" question as well. The current model presumably could also be used prescriptively. Through a series of device designs and modifications, several $\underline{\tau}$s would be generated. The process of device redesign would be continued until $\underline{\tau}$ reached a sufficiently high level. However, this would obviously be a very inefficient procedure.

An alternative would be to expand the model to include a specific "prescriptive mode" of analysis. This expansion would aid the device designer in identifying device improvements which would have a large and positive impact on effectiveness. These improvements could be based on considerations already included in the model, such as the importance of high deficit subtasks, functional similarity, and so on. However, this expansion would not be straightforward. For example, simply increasing the physical similarity of a device might actually <u>reduce</u> effectiveness if functional similarity were not improved correspondingly. Thus, this model extension would have to address fundamental issues of the interaction among the model's parameters.

4.4 Conclusions and Recommendations

Traditionally the Army has relied upon operational hardware to provide skill training to officers and enlisted personnel during assignment to field units. Increasingly, however, this practice has been giving way to the use of simulated training devices. A number of reasons for this shift in training philosophy have been advanced, including policy considerations as well as considerations having to do with training technology per se. It has become axiomatic among educational/training specialists that the complex processes of learning are not necessarily best served by "hands on" experience with real equipment. Instead, these processes may be

better served by the simulative device, since it, unlike the operational equipment, can be specifically designed and employed to optimize a variety of instructional features.

With the advent of the system engineering approach to the design of instruction, the simulator has become potentially even more important. It lends itself to the system approach particularly well and in ways not feasible with operational equipment. Thus, for a variety of reasons, a trend has been established toward replacing operational equipment with equipment simulators in order to develop and maintain the skills of Army personnel. As compelling as these reasons are, there are a number of countervailing factors which are equally compelling and which require sober consideration.

The first of these factors is the cost of developing and producing simulators. This cost may be considerably more than the actual equipment because of the inclusion of instructional features, and the more flexible the simulator is with respect to these features, the greater the expense. The second factor is limited knowledge about the role of a variety of potent variables and instructional features in promoting transfer of training. In fact, capacity for building training system components, including sophisticated training simulators, has far outstripped knowledge about how to design them for effective training, and how and when to use them vis-a-vis the specific behavioral objectives to be achieved.

Thus, while the need for increasing use of training simulators is clear cut, it is equally evident that their cost-effectiveness and transfer efficiency cannot be taken for granted. *Some means must be found for* evaluating training simulators and for doing so within a broad system context that includes all the classes of variables which may promote or limit training device effectiveness. Ideally, this evaluation should be applicable during early stages of the device development cycle so that alternative device designs can be contrasted and potential effectiveness predicted.

The current research program has addressed itself to this general problem. More specifically, *a conceptual framework or* model has been developed which provides guidelines for predicting and evaluating the effectiveness of training devices under development. The model takes into consideration

what must be trained, who must be trained, and how the device is to be used. Evaluations of the model have established its feasibility and reliability of application and have provided provisional support for its predictive validity.

As suggested by the potential revisions and expansions discussed above, however, much research remains to be done. Additional evaluations of the model's predictive validity will be required in the future to insure its utility as a means for forecasting effectiveness. Similar research will be needed to establish the construct validity of its parameters and to determine how those parameters interact to influence estimates of effectiveness. Equally important, steps must be taken to ease the model into the Army's design and development cycle. Toward this end, expansions of the model should be considered which address device effectiveness in terms of savings measures and later transfer. Simultaneously, work should focus on development of procedures implicit in the model which device designers and developers can use to prescribe effective training solutions.

Even without such embellishments the model as it now stands can serve as a useful tool in the design and development process. Its chief value lies in formally directing attention to the important issues which should be considered during device design, development, and evaluation. It makes explicit the need to consider the interactive effects of different kinds of variables on training effectiveness, and provides a formal way of pursuing the system engineering approach to Army training.

REFERENCES

Adams, J. A. Psychomotor response acquisition and transfer as a function of control indicator relationships. Journal of Experimental Psychology, 1954, 48(1), 10-18.

Bernstein, B. R., & Gonzalez, B. K. Learning, retention and transfer in military training. NAVTRADEVCEN 69-C-0253-1, Contract N61339-69-C-0253, Honeywell, Inc., 1971, AD733964.

Blaiwes, A. S., & Regan, J. J. An integrated approach to the study of learning, retention and transfer--a key issue in training device research and development. NAVTRADEVCEN IH-178, Naval Training Device Center, 1970, AD712096.

Chenzoff, A. P., & Folley, J. D., Jr. Guidelines for training situation analysis (TSA). Applied Science Associates, Inc., Valencia, Pennsylvania, 1965, AD472155.

Continental Army Command. Training. Systems Engineering of Training Course Design. CON REG 350-100-1, April, 1972.

Continental Army Command. Training. Systems Engineering of Unit Training. CON PAM 350-11, January, 1973.

Demaree, R. G. Development of training equipment planning information. USAF ASD Technical Report 61-533, October, 1961, AD267236.

Fitts, P. M., & Posner, M. I. Human Performance. Belmont: Brooks-Cole, 1967.

Folley, J. D., Jr. Development of an improved method of task analysis and beginnings of a theory of training. Applied Science Associates, Inc., Valencia, Pennsylvania, 1964, AD445869.

Gagné, R. M., Foster, H.; & Crowley, M. E. The measurement of transfer of training. Psychological Bulletin, 1948, 45, 97-130.

Osgood, C. E. The similarity paradox in human learning: A resolution. Psychological Review, 1949, 56, 132-143.

Pew, R. W. The speed-accuracy operating characteristic. Acta Psychologica, 1969, 30, 16-26.

Rose, A. M., Wheaton, G. R., and Leonard, R. L., Jr., Fingerman, P. W., and Boycan, G. G. Evaluation of two tank gunnery trainers. ARI Research Memorandum 76-19, August 1976.

U.S. Naval Training Device Center. Staff study on cost and training effectiveness of proposed training systems. TAEG Report 1, U.S. Naval Training Device Center, Orlando, Florida, 1972.

Wheaton, G. R., Rose, A. M., Fingerman, P. W., Korotkin, A. L., and Holding, D. H. Evaluation of the effectiveness of training devices: Literature review and preliminary model. ARI Research Memorandum 76-6. April 1976.

Wheaton, G. R., Rose, A. M., Fingerman, P. W., Korotkin, A. L., and Holding, D. H. Evaluation of the effectiveness of training devices: II. Evaluation plan for preliminary model. Second Interim Report. U. S. Army Research Institute Contract No. DAHC 19-73-0049. Washington, D. C.: American Institutes for Research, 1974(b).

Wheaton G. R., Fingerman P. W., Rose, A. M., Leonard, R. L., Jr. Evaluation of the effectiveness of training devices: Elaboration and application of the predictive model. ARI Research Memorandum 76-16. July 1976.

Wheaton, G. R., Rose, A. M., Fingerman, P. W., Leonard, R. L., Jr., and Boycan, G. G. Evaluation of tnree burst-on-target trainers. ARI Research Memorandum 76-18. August 1976.

APPENDIX I

Procedures for Application of Predictive Model

Introduction

Procedures for using the predictive model in order to obtain estimates of training device effectiveness were detailed in an earlier report in the current series (Wheaton, Fingerman, Rose, & Leonard, 1976). That presentation was complicated, however, by discussion of the rationale underlying components of the predictive model, of the logic behind the analytical steps required for its application, and of the results obtained during initial evaluations. Much of this material is of secondary interest to persons simply wishing to apply the model to a training device.

Accordingly, this Appendix has been prepared in the form of a "cookbook," providing the step-by-step procedures involved in generating an effectiveness prediction. The pr dure has been broken down into three parts, including: 1) preparation of input data; 2) application of model; and 3) generation of effectiveness estimates. Examples are provided throughout for three BOT devices used to train students to apply BOT on stationary targets using the 3A102B laser device. (Details concerning the devices and transfer task are presented in an earlier report by Wheaton, Rose, Fingerman, Leonard, & Boycan, 1976.)

As will become apparent throughout th cussion of the model's application, the person performing the following analyses must be knowledgeable and sophisticated. We have attempted to eliminate arbitrariness as far as possible; however, skilled judgments are still necessary at several points in the analyses. Device evaluation still is an art as well as a skill.

Preparation of Input Data

There are three kinds of information which must be obtained before application of the model can proceed: 1) a detailed statement of the training objective; 2) a task analysis of the operational (transfer) task including a list of the controls and displays and a list of the skills and knowledges for each subtask in the transfer task; and 3) a task analysis of the device

(training) task including a display and control list for each subtask in the device task. Each of these requirements is discussed below.

Statement of Training Objective. The most basic data requirement is a detailed statement of the training objective to include: 1) the precise nature of the task to be learned (i.e., to be transferred to); 2) the conditions of task performance during transfer; and 3) the performance standard(s) to be met. The procedures for developing a detailed statement of the training objective have been formalized and are presented in CON REG 350-100-1 (1972) and CON PAM 350-11 (1973).

Task Analysis of Transfer Task. A second major data requirement is detailed information regarding the operational criterion (transfer) task specified in the training objective. Identification and listing of the subtasks comprising the operational criterion task is essential.

In order to generate such data the general approach described in CON PAM 350-11 (1973) and the specific guidelines provided by Folley (1964) and Chenzoff and Folley (1965) may be used. The approach is to break the operational criterion situation down into successively finer units of description, stopping at what constitutes the subtask level of detail. Based on Folley's (1964) system of description, a subtask may be defined as an activity that is performed by one person and bounded by two events. An example of a subtask might be "Upon receipt of the alert element of a fire command, sets turret power switch to the 'ON' position." An event may be defined as a discrete and identifiable act or occurrence. Examples would be, "receipt of alert element" and "switch in 'ON' position." An activity is defined as the behavior(s) comprising a subtask, such as "setting a switch." A task is defined as a set of two or more subtasks (e.g., "fire main gun using the M-32 sight").

As an example, eight subtasks were identified as comprising a transfer task in which the 3A102B laser was used to apply BOT to static targets. The eight subtasks are listed down the left-hand margin of Tables 1 and 2. Once the relevant subtasks have been identified and listed, they are analyzed to detail further the precise activities involved in the performance

of each subtask, to identify the displays and controls which the operator utilizes, and to determine the skills and knowledges involved.

As shown in Table 1, the displays (D) and controls (C) relevant to each subtask are listed under that subtask along the left margin. A display is defined as an information source or transmitter, and a control as an information receiver which must be physically operated on. Information is defined in the information theoretic sense used by Fitts and Posner (1967). A display (D) or control (C) is included in the list generated for each subtask if it either receives or transmits the information involved in performance of the subtask. In subtask 3, for example, information is transmitted by the sight reticle and is received (from the operator) by the cadillac tracking controls.

As shown in Table 2, the skills (S) and knowledges (K) related to performance of each subtask are also listed. Based on the task descriptive data which provides information about actual performance of the subtask on the operational equipment and on the training objective statement, a sentence is written which describes each activity in the subtask. From this statement knowledges and skills are inferred and listed. The distinction between skills and knowledges is not critical and is only made for the convenience of the analyst. Current Army task-analytic methods provide for a listing of basic skills and knowledges, but these should be augmented by focusing on the detailed subtasks.

Task Analysis of Training (Device) Task. The third and final data requirement is a detailed task analysis of the training task on which students will practice and from which they will transfer to the operational task. Analysis of the training task should parallel that described above for the operational task with one exception. Lists of skills and knowledges are not required. The results of this analysis will be a list of subtasks, as well as a list of the specific displays and controls involved in the performance of each subtask.

Application of Model

Given the inputs described above one can begin to apply the predictive model. Application consists of deriving values for five parameters within

the model: 1) communality, 2) physical similarity, 3) functional similarity, 4) learning deficit, and 5) training techniques. The procedures which must be followed in generating values for each of these parameters are described below.

Task Communality. The first step is to construct a *task communality* matrix as shown in Table 1. Subtasks are listed down the left margin, and the training devices to be assessed are listed across the top of the page (under the communality analysis heading). The second step is to list the subtasks comprising the training task. This listing is accomplished separately for each device under consideration. For accuracy and to insure reliability it is suggested that this step be carried out formally. Potentially valuable information may be overlooked if one simply considers each operational subtask and makes a guess about its inclusion in the device.

Armed with lists of subtasks for the device and operational setting, one can perform the third and crucial step. For each operational subtask listed along the left margin, the analyst scans his list of training subtasks. If, in fact, a device enables the trainee to practice the subtask in question, a "1" is entered in the appropriate cell under the device. However, if that particular subtask is not represented, a "0" is entered. This process is continued until all operational subtasks have been evaluated. The issue of how well or how faithfully the subtask is represented is dealt with during Similarity analysis. In some cases there will be additional subtasks associated uniquely with a device and not found in the operational setting. These subtasks should be footnoted at the bottom of the task-communality matrix, and retained for further analysis. In the example shown in Table 1 each of the eight criterion task subtasks is represented in the three training devices as indicated by the "1" entries.

Physical Similarity Analysis. The second step in applying the model is to conduct a Physical Similarity analysis. The assessment is based on the physical similarity or fidelity of displays and controls in the training device relative to those in the operational equipment. For each subtask, a rating is made on each relevant operational control and display which describes how well each is represented in the training device. While

ratings of subtasks lacking in communality are not used directly, it is generally useful to make ratings for all subtasks. The ratings of physical similarity are made along the following four-point scale:

Rating	Definition
3	Identical. The trainee would not notice a difference between the training device control or display and the operational control or display at the time of transfer. Note that they need not be absolutely identical, but there must be no "jnd" (just noticeable difference) for the trainee. Include for consideration the location, appearance, feel, and any other physical characteristics. Ignore the amount and quality of information transmitted.
2	Similar. There would be a jnd for the trainee at the time of transfer, but he would be able to perform the task. There might be a decrement in performance at transfer, but any such decrement would be readily overcome.
1	Dissimilar There would be a large noticeable difference, quite apparent to the trainee, at transfer, and a large performance decrement, given that the trainee could perform at all. Specific instruction and practice would be required on the operational equipment after transfer to overcome the decrement.
0	The control or display is not represented at all in the training device.

As shown in Table 1, the ratings for each equipment component are entered in the appropriate cell corresponding to an operational subtask equipment component and a particular training device. Summary physical proficiency scores may be calculated for each subtask by a weighted averaging procedure. For example, physical similarity for the 17-4 device and subtask 3 is obtained by dividing the sum of the ratings for that subtask (9) by the number of displays and controls rated (6). The obtained value (1.5) is then divided by 3 to scale physical similarity between zero and one (in

this case, .5). Summary data are recorded in the appropriate columns as shown in Table 1.

Functional Similarity Analysis. The next step in applying the model is to conduct a Functional Similarity analysis. This analysis examines the operator's behavior in terms of the information flow from each display to the operator, and from the operator to each control. The assessment is made in terms of the amount of information transmitted (Fitts & Posner, 1967) from each display to each control (regardless of the actual operational mode of transmission or reception), and the type of information-processing activity performed by the operator. Thus, regardless of the physical characteristics of a control or display, the issue is whether the operator acts upon the same amount of information in the same way in both the operational and training situations.

The analysis makes use of the subtask descriptions and the list of the controls and displays in the operational situation as shown in Table 1. Controls and displays are considered in conjunction with subtask activities to determine the type, amount, and direction of information flow within each subtask. Each situation in which a display transmits information to the operator (e.g., the operator reads the display) is defined as a stimulus function, and each situation in which the operator transmits information to a control (e.g., operates it) is termed a response function. Thus, the derived input for the Functional Similarity analysis is the list of information-processing functions indicated by the controls and displays of the operational situation.

In each subtask, the number of bits of information is determined for each stimulus and response situation, by estimating the number of states which the display or control may assume. The amount of information in the operational setting (H_{os}) is equal to $\log_2$ of the number of states in the stimulus or response functions under consideration. The amount of information in the training setting (H_{ts}) for each of the corresponding functions is estimated in the same way for each training device. Each stimulus and response function is then rated according to the following four-point scale:

Rating	Definition
3	Identical. $H_{ts} = H_{os}$.
2	Similar. $H_{ts} \sim H_{os}$; they are within one $\log_2$ unit of each other.
1	Dissimilar. $H_{ts} \neq H_{os}$; they are more than one $\log_2$ unit apart.
0	Missing. $H_{os} > 0$, and $H_{ts} = 0$.

In certain cases, ratings of 2 and 1 will be assigned to situations that have been purposely made unequal by the device designer in order to implement some training technique (e.g., augmented feedback or guidance). Such cases should be footnoted for consideration in the Training Techniques stage of the analysis. In other cases ratings of 3 will be assigned when the amount of information is the same or nearly so, but when the form of the information is radically different. For example, in the operational task the operator might index ammunition by pulling and turning the index handle. This handle could assume 6 positions; therefore, indexing ammunition is a $\log_2$ 6-bit task. In a training device, the same six alternatives might be present; however, ammunition might instead be indexed by pressing one of six buttons. The trainee might process this same information in a completely different way, or use a different strategy to deal with it. Such cases should also be footnoted for later consideration.

Ratings for the three training devices are shown in Table 1 for the stimulus and response functions (displays and controls) associated with each operational subtask. Summary values are obtained for each subtask following the same procedure described above for physical similarity. These data are then entered in the appropriate columns as shown in Table 1.

Learning Deficit Analysis. This step actually involves three separate steps which are designed to: 1) assess the skills and knowledges in the student's repertory before training, 2) determine the skills and knowledges which he must possess at the time of transfer to the operational setting, and 3) estimate the difficulty (in terms of training time) of training

the necessary skills and knowledges. The output of this stage of the analysis is a number for each subtask indicating the deficit possessed by the typical trainee, weighted by the relative difficulty (in terms of estimated training time on the operational equipment) of surmounting that deficit.

The analysis begins with the application of a rating scale to estimate the "amount" of each skill or knowledge which the average trainee (of the type selected for course enrollment) could be expected to have upon his first exposure to the training system or device. The following Repertory Scale (RS), adapted from Demaree (1961), is used to describe the level of each skill and knowledge in the student's repertory prior to the start of formal training:

Rating	Definition
0	No experience, training, familiarity, etc., with this skill or knowledge. Cannot perform a task requiring this skill or knowledge.
1	Has only a limited knowledge of this subject or skill. Has not actually used the information or skill. Cannot be expected to perform. Has had "orientation" only.
2	Has received a complete briefing on the subject or skill. Can use the knowledge or skill only if assisted in every step of the operation. Requires much more training and experience. Has received "familiarization" training only.
3	Understands the subject or skill to be performed. Has applied part of the knowledge or skill either on the actual job or a trainer. Has done the job enough times to make sure he can do it, although perhaps only with close supervision. Needs more practice under supervision. Has had "procedural" training.
4	Has a complete understanding of the subject or skill. Can do the task completely and accurately without supervision. Has received "skill" training.

Obtained ratings for the knowledges and skills underlying each subtask in the transfer task are shown in Table 2 in the "RS" column.

After the analyst has assessed the level of skills and knowledges in the trainee's repertory, he proceeds to determine the required "amount" of each skill or knowledge which the trainee must possess at the close of training and time of transfer. The following Criterion Scale (CS), adapted from Demaree (1961) is used:

Rating	Definition
0	At the end of training, the trainee should have no experience or training.
1	Should have a limited knowledge of the subject or skill. Has not actually used the information. Is not expected to perform the task. Has completed "orientation" training.
2	Should have received a complete briefing on the subject or task. Is able to use the knowledge or skill only if assisted in every step of the operation. Requires much more training and experience to be able to perform the task independently. Has had "familiarization" training.
3	Should have an understanding of the subject or skill to be performed. Has applied part of the knowledge or skill on the actual job or a trainer. Has done the job enough times to make sure he can do it, although perhaps only with close supervision. Needs more practice under supervision. Has had "procedural" training.
4	Should have a complete understanding of the subject, or be highly skilled. Is able to perform the task completely, accurately, and independently. Has had "skill" training.

Obtained "CS" ratings are shown in Table 2.

The next step is to calculate the learning deficit by subtracting the repertory rating (RS) from the criterion rating (CS) for the knowledges and skills underlying each subtask. Negative differences are set equal to zero, since they indicate that the trainee enters training with more proficiency than is necessary, and so has no deficit. The difference scores are then averaged within each subtask to obtain a mean subtask deficit as shown in Table 2.

The next step is to rank the subtasks in terms of estimated training time, assuming that only the operational equipment would be available for training. The analyst begins by seeking out the subtask which would require the least training time on the operational equipment, and assigns it a rank of "1". The subtask requiring the next smallest amount of training time for surmounting its associated deficit is assigned a rank of "2", and so on, until all subtasks have been ranked. The ranks obtained for the subtasks in our example are shown in Table 2. Next, the mean subtask deficits are multiplied by their corresponding ranks, to obtain a weighted learning deficit score. Finally, each such score is divided by 4 times the number of subtasks, to provide an index between 0 and 1 which reflects the size and importance of the deficit on each subtask relative to the other subtasks being analyzed. The obtained values are entered as indicated in Table 2.

Training Technique Analysis. As the fifth step in applying the model, a Training Technique analysis is conducted which relates the particular skills and knowledges which must be trained for each subtask to a set of principles and techniques which describe how best to train various kinds of content. Two steps are involved. Subtasks are described in terms of taxonomic categories and judgments are made of the extent to which training principles relevant to those categories have been incorporated in the training device.

The first step is to assign the subtasks to taxonomic categories using the following labels (after U.S. Naval Training Device Center, 1972).

1. Recalling facts and principles
2. Recalling procedures
3. Nonverbal identification
4. Nonverbal detection
5. Using principles, interpreting, inferring
6. Making decisions
7. Continuous movement
8. Verbal detection and identification
9. Positioning and serial movement

10. Repetitive movement
11. Written verbalization
12. Oral verbalization
13. Other verbalization, including signs

It is not necessary to restrict subtasks to a single category; multiple labels are permissible. The categories used to describe the eight exemplary subtasks are listed in Table 2 according to the number code used above. In several cases, more than one taxonomic label has been applied to a subtask.

The second and major step is for the analyst to evaluate the training principles/techniques listed in Appendix I of the 1976 report by Wheaton, Fingerman, Rose, and Leonard.

The training techniques are organized along two independent dimensions. First, they have been coded according to the taxonomic category to which they apply. Second, within each taxonomic category, they have been further organized into techniques relevant to stimulus considerations, response considerations, or feedback considerations. Thus, by referring to the taxonomic label(s) which he has assigned to each subtask, the analyst can draw out those principles/techniques which correspond to the set of relevant behavioral categories, and sort them into three groups: stimulus, response, and feedback. With the operational task information and the training device and system information before him, he rates the training device for each relevant principle in each of the three categories. While performing the rating operation, he should pay special attention to any items from previous portions of the analysis which were "flagged" for attention at this stage. The ratings are made from the following scale:

Rating	Definition
3	Optimal implementation of this technique; in complete accord with this principle.
2	Good implementation of this technique; in excellent accord with this principle.
1	Fair implementation of this technique; good accord with this principle.

Rating	Definition
0	This principle or technique was inapplicable or irrelevant. OR The device neither implemented this technique nor violated this principle.
-1	Mild violation of this training principle; implementation of a mildly opposing technique.
-2	Serious violation of this principle or technique.
-3	Complete violation of this principle; implementation of a strongly contraindicated technique.

For each subtask, the lowest obtained rating for each of the stimulus, response, and feedback considerations is selected and recorded as shown in Table 2. These three ratings are then averaged, the constant 3 is added to the obtained mean (to delete negative signs), and this sum is divided by 6 to provide an index between 0 and 1 yielding the training technique score of the training device *for each subtask.* These values have been entered, for example, in the three columns along the right margin of Table 2. Application of the model is complete once this last step has been accomplished. Given values for the parameters described above, one can proceed to the generation of effectiveness estimates.

Generation of Effectiveness Estimates

Once parameter values have been derived for a training device, as described above, one can generate estimates of the training effectiveness of the device. Two steps are involved, including development of a parameter summary table, and insertion of the parameter values into an effectiveness equation.

In preparing the summary table one begins by listing the operational subtasks along the left margin as illustrated in Table 3. The list has been presented three times in Table 3 corresponding to the three training devices under evaluation. Subtask communality estimates are then entered (from Table 1). Similarity estimates representing a pooling of physical and functional similarities are obtained by averaging these two parameters

for each subtask. The average similarity (S) is then recorded as shown in Table 3. Mean deficit scores, and subtask rank order learning difficulty are also entered (from Table 2). Values for these last two parameters are further processed to yield yet another parameter ÷ a weighted learning difficulty (WLD) score. For each subtask, the mean deficit is multiplied by the subtask rank difficulty and the product is divided by four times the number of subtasks listed. The result is listed in the WLD column. Finally subtask values for training techniques are entered (from Table 2).

Effectiveness estimates are obtained in the following manner. For each subtask, the values for C, S, WLD, and T are multiplied together. The resulting products are then added together to obtain a sum for the overall task as represented in a specific training device. For example, using data shown in Table 3 the sum for the 17-4 device would be equal to .300 (i.e., .015 + .021 + .037 + .012 + .065 + .109 + .041 + .000). This sum is in turn divided by the sum of the WLD's (1.13) to yield an effectiveness estimate of .265 for the 17-4 device.

This estimate is bounded by 0 and 1. The higher the value, the greater is the predicted training effectiveness for the device in question. As the value decreases toward zero, effectiveness decreases, approximating the performance expected if an untrained group of students first performing the operational task.

TABLE I-1

Communality and Similarity Analyses

Subtasks in BOT Transfer Task	COMMUNALITY ANALYSIS			SIMILARITY ANALYSIS Physical			Functional			Physical Summary Similarity			Functional Summary Similarity		
	17-4	17-4M	17-B4	17-4	17-4M	17-B4	17-4	17-4M	17-B4	17-4	17-4M	17-B4	17-4	17-4M	17-B4
1. Receives Elements of Fire Command	1	1	1							1	1	1	1	1	1
D_1 Ears (Auditory Message)				3	3	3	3	3	3						
2. Monitors for & Identifies Target	1	1	1							.78	.78	.89	.89	.89	1
D_1 M32 Sight				2	2	3	2	2	3						
D_2 Target (Square)				2	2	2	3	3	3						
C_1 Voice ("Identified")				3	3	3	3	3	3						
3. Aims	1	1	1							.50	.50	.94	.67	.67	.83
D_1 Target (Square)				2	2	2	2	2	2						
D_2 Reticle				2	2	3	3	3	3						
D_3 M32 Sight				2	2	3	2	2	3						
D_4 Ears ("Fire" & "Up")				3	3	3	3	3	3						
C_1 Palm Switches				0	0	3	0	0	3						
C_2 Cadillac Controls				0	0	3	2	2	1						
4. Fires	1	1	1							.83	.83	1	1	1	1
C_1 Voice ("On the way")				3	3	3	3	3	3						
C_2 Trigger Switch				2	2	3	3	3	3						
5. Senses the Round	1	1	1							.67	.67	.89	.78	.78	1
D_1 M32 Sight				2	2	3	2	2	3						
D_2 Laser Burst				2	2	3	2	2	3						
D_3 Target				2	2	2	3	3	3						
6. Determines New Aiming Point	1	1	1							.67	.67	1	.67	.67	1
D_1 M32 Sight				2	2	3	2	2	3						
D_2 Laser Burst				2	2	3	2	2	3						
D_3 Reticle				2	2	3	2	2	3						
7. Re-Aims	1	1	1							.50	.50	.94	.61	.61	.83
D_1 M32 Sight				2	2	3	2	2	3						
D_2 Target				2	2	2	2	2	2						
D_3 Reticle				2	2	3	2	2	3						
D_4 Ears ("Up")				3	3	3	3	3	3						
C_1 Palm Switches				0	0	3	0	0	3						
C_2 Cadillac Controls				0	0	3	2	2	1						
8. Fires	1	1	1							.83	.83	1	1	1	1
C_1 Voice ("On the way")				3	3	3	3	3	3						
C_2 Trigger Switches				2	2	3	3	3	3						

TABLE I-2

Learning Deficit and Training Technique Analyses

Subtasks in BOT Transfer Task	RS	CS	CS-RS	MEAN CS-RS	RANK DIFFICULTY	WLD	TAXONOMIC LABELS	STIMULUS 17-4	STIMULUS 17-4M	STIMULUS 17-B4	RESPONSE 17-4	RESPONSE 17-4M	RESPONSE 17-B4	FEEDBACK 17-4	FEEDBACK 17-4M	FEEDBACK 17-B4	SUMMARY TRAINING TECHNIQUE ANALYSIS 17-4	SUMMARY TRAINING TECHNIQUE ANALYSIS 17-4M	SUMMARY TRAINING TECHNIQUE ANALYSIS 17-B4
1. Receives Elements of Fire Command				1.00	1	.03	1,2,8	0	0	0	0	0	0	0	0	0	.5	.5	.5
K1: Know elements of fire command (cue)	3	4	1																
2. Monitors for & Identifies Target				.40	4	.05	1,2,3,4,12	0	0	-1	0	0	0	0	0	0	.5	.5	.44
S1: Operate M32	3	4	1																
S2: Identify announced target	4	4	0																
K2: Know relevant targets	3	4	1																
K3: Knows to announce "Identified"	4	3	0																
K4: Knows location of M32	4	4	0																
3. Aims				1.20	6	.23	1,2,9	-1	-1	-1	-2	-1	-3	-1	-1	-3	.28	.33	.11
S3: Operate M32	3	4	1																
S4: Operate turret controls	2	4	2																
S5: Must aim accurately	2	4	2																
K5: Knows location of M32 & cadillac controls	4	4	0																
K6: Know where to and how to aim	3	4	1																
4. Fires				.33	2.5	.03	2,9,12	0	0	0	-1	-1	0	0	0	0	.44	.44	.5
K7: Know location of trigger	4	4	0																
K8: Know firing procedure	3	3	0																
K9: Know to monitor for strike	3	4	1																
5. Senses the Round				1.00	5	.16	2,4	0	0	0	0	0	0	1	1	0	.56	.56	.5
S6: Operate M32	3	4	1																
S7: Sensing round vis a vis target	2	4	2																
K10: Know position of M32 sight	4	4	0																
K11: Know sensing procedure	3	4	1																
6. Determines New Aiming Point				1.50	8	.37	1,2,3,5	-1	-1	0	0	0	0	0	0	0	.44	.44	.5
S8: Operate M32	3	4	1																
S9: Determine burst vis a vis reticle	1	4	3																
K12: Know new aiming point procedure	2	4	2																
K13: Know location of M32	4	4	0																
7. Re-Aims				1.20	7	.26	5,9	-1	-1	-1	-2	[illegible]	[illegible]	[illegible]	-1	-3	.28	.33	.11
S10: Operate M32	3	4	1																
S11: Operate cadillac controls	2	4	2																
S12: Place new aim point on target	2	4	2																
K14: Know location of M32 & controls	4	4	0																
K15: Know re-aiming procedure	3	4	1																
8. Fires				.00	2.5	.00	2,9,12	0	0	0	-1	-1	0	0	0	0	.44	.44	.5
K16: Know location of trigger	4	4	0																
K17: Know refire procedure	3	3	0																

TABLE I-3

PARAMETER SUMMARY TABLE

17-4

BOT TASK	Communality (C)	Similarity (Mean)	Mean Learning Deficit (LD)	Rank Learning Difficulty (R)	Weighted Learning Difficulty (WLD)	Training Technique Analysis (T)
1. Alert	1	1	1	1	.03	.5
2. Identify	1	.83	.4	4	.05	.5
3. Aim	1	.59	1.2	6	.23	.28
4. Fire	1	.91	.33	2.5	.03	.44
5. Sense	1	.73	1	5	.16	.56
6. Apply BOT	1	.67	1.5	8	.37	.44
7. Reaim	1	.55	1.2	7	.26	.28
8. Fire	1	.91	0	2.5	0	.44

17-4M

BOT TASK	Communality (C)	Similarity (Mean)	Mean Learning Deficit (LD)	Rank Learning Difficulty (R)	Weighted Learning Difficulty (WLD)	Training Technique Analysis (T)
1. Alert	1	1	1	1	.03	.5
2. Identify	1	.83	.4	4	.05	.5
3. Aim	1	.59	1.2	6	.23	.33
4. Fire	1	.91	.33	2.5	.03	.44
5. Sense	1	.73	1	5	.16	.56
6. Apply BOT	1	.67	1.5	8	.37	.44
7. Reaim	1	.55	1.2	7	.26	.33
8. Fire	1	.91	0	2.5	0	.44

17-B4

BOT TASK	Communality (C)	Similarity (Mean)	Mean Learning Deficit (LD)	Rank Learning Difficulty (R)	Weighted Learning Difficulty (WLD)	Training Technique Analysis (T)
1. Alert	1	1	1	1	.03	.5
2. Identify	1	.95	.4	4	.05	.44
3. Aim	1	.89	1.2	6	.23	.11
4. Fire	1	1	.33	2.5	.03	.5
5. Sense	1	.95	1	5	.16	.5
6. Apply BOT	1	1	1.5	8	.37	.5
7. Reaim	1	.89	1.2	7	.26	.11
8. Fire	1	1	0	2.5	0	.5